



Attwood, A., & Munafo, M. (2016). Navigating an open road. *Journal of Clinical Epidemiology*, 70, 264–266. DOI: [10.1016/j.jclinepi.2015.04.016](https://doi.org/10.1016/j.jclinepi.2015.04.016)

Peer reviewed version

Link to published version (if available):
[10.1016/j.jclinepi.2015.04.016](https://doi.org/10.1016/j.jclinepi.2015.04.016)

[Link to publication record in Explore Bristol Research](#)
PDF-document

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/pure/about/ebr-terms.html>

Navigating an Open Road

Angela S. Attwood ^{1,2} and Marcus R. Munafò ^{1,2}

1. MRC Integrative Epidemiology Unit at the University of Bristol, United Kingdom.
2. UK Centre for Tobacco and Alcohol Studies, School of Experimental Psychology, University of Bristol, United Kingdom.

Corresponding author: Marcus R. Munafò, School of Experimental Psychology, University of Bristol, 12a Priory Road, Bristol BS8 1TU, United Kingdom. T: +44.117.9546841; F: +44.117.9288588; E: marcus.munaf@bristol.ac.uk

Keywords: Open Science; Open Materials; Open Data; Open Access; Reproducibility.

Navigating an Open Road

West [1] outlines an agenda for promoting greater transparency and accountability through the routine disclosure of data and associated materials such as statistical commands. The Tobacco and Alcohol Research Group (TARG; <http://www.bris.ac.uk/expsych/research/brain/targ/>), part of the UK Centre for Tobacco and Alcohol Studies and the MRC Integrative Epidemiology Unit at the University of Bristol, has been moving towards an (admittedly incomplete) Open Science model over the last few years, focused on three core areas: materials (specifically study protocols), data and publications. Here we discuss our experience, and highlight what we consider to be the main advantages, as well as potential pitfalls and cases where full openness may not be achievable.

Open Science is an umbrella term encompassing a movement that encourages scientists to make their materials, data and publications freely available to all. In its broadest sense it includes open source software, open peer-review (such as that practiced by the *Frontiers* family of journals) [2], and other resources (such as educational materials). West focuses his argument on the benefits of Open Data (taken here to also include the associated statistical commands): it should serve to reduce the error rate, and facilitate additional analyses. We agree, but argue that the same principle of openness can and should (where possible) be extended to other aspects of the scientific process. It is a substantial undertaking, but one which (in our experience) can generate unexpected benefits.

Open Materials

The 2013 revision of the Declaration of Helsinki recommended, for the first time, that all research studies using human participants should be publicly registered prior to data collection. There are a number of potential benefits associated with the pre-registration of study protocols. First, it should improve the standard of published

research by preventing post-hoc adjustment of study aims and hypotheses, and deviations from planned statistical analyses. Second, it serves to clearly differentiate between exploratory analyses that pursue interesting and unexpected findings from those that were pre-planned. Third, it reduces the likelihood that ongoing research efforts will be duplicated (although one potential concern is that others may take advantage of openness to pursue similar ideas themselves). Fourth, if study protocols are peer-reviewed (e.g., through a specific submission format such as Registered Reports [3]), researchers may obtain detailed feedback on their study design, methods and analysis plans.

A number of medical journals, such as *Trials*, *BMJ Open*, and the *BMC* family of journals, offer peer-reviewed publication of study protocols, but similar mechanisms are less common in other fields. In 2013, the journal *Cortex* announced a new publication initiative, involving a two-stage peer-review process in which authors first submit a report detailing experimental methods and planned analyses prior to data collection [3]. This ensures that the initial peer-review focuses on methodological quality and is not influenced by the results. It also serves to protect against the implicit pressure to “find” interesting results, since the acceptance-in-principle offered after first stage review guarantees eventual publication as long as the authors adhere to the original study protocol. Variations on the Registered Reports format are now offered by a number of journals [3-7].

An alternative approach is to use online platforms such as the Open Science Framework (OSF: <https://osf.io/>), which enable researchers to make protocols publicly available and provide a date-stamped digital object identifier (DOI) that can be cited in future publications. We have used the OSF platform since early 2014, and initially experienced difficulties that were due to, or exacerbated by, unfamiliarity with the system (e.g., not recognizing the point at which the system “locked”, which meant that some protocols were posted with minor elements missing or incomplete). However, once mastered, the process is straightforward, and we have been able to

extend this to all relevant research activity involving primary data collection. We also worked to ensure that the formatting and content of our protocols was standardized as much as possible.

Our experience has been that the process of pre-registering study protocols, either via publication or an online platform, encourages careful consideration of all aspects of the research, from study design through to the dissemination of findings, and acts as a useful reminder regarding what was hypothesized and what analysis plan was decided upon (since analysis may take place months or even years after the study was originally conceived). The main challenge we have faced is the additional time required to format, proofread and finalize protocols; however, much of this time is recouped later since much of the work required to prepare a manuscript for publication has already been completed.

Open Data

Open Data is a relatively new phrase, but the concept is not. It refers to the idea that research data should be made publicly available free of charge and, where possible, free to use without sharing restrictions. The main arguments for Open Data include promoting transparency and maximizing the return on research funding through the wider use of datasets. In addition, since much research is publicly funded, there is strong argument for it to be freely available to the general public. An increasing number of peer-reviewed journals are making Open Data a requirement of publication, such as the PLOS family of journals [8].

However, there are a number of issues that may make Open Data difficult to achieve. First, data may have been collected without explicitly obtaining consent for the resulting data to be made publicly available. This is particularly likely in historical or ongoing studies (e.g., large cohort studies), where ethics approvals and associated informed consent procedures may have been put in place before Open Data was widely mandated or encouraged. Second, it may be theoretically possible

to identify individuals (or for individuals to self-identify themselves) from information included in a dataset, particularly in small studies of rare conditions conducted in a limited geographical area. This is most sensitive if some of the information included in the dataset could potentially be misleading or distressing when taken in isolation and without appropriate advice or information.

As part of our move towards an Open Science model, we therefore had to revise our ethics submissions and informed consent procedures so that participants now provide explicit consent for their data to be made public. We avoid statements indicating that data will only be used for *research* purposes, since it is not possible to guarantee this within an Open Data model. This process required detailed discussions with our institutional ethics committee and research governance office. During these discussions, we were asked to consider adding an "opt-out" clause to Open Data on our consent forms. This is an interesting point and there is obvious merit in giving our research participants control over their data if these are shared. However, it is problematic since in this situation publicly available datasets may not contain the full dataset on which reported research findings are based. This would undermine one of the fundamental aims of Open Data, and we therefore did not adopt this suggestion.

There are a growing number of research data repositories available for archiving Open Data. Some are subject-specific but many support Open Data from a variety of disciplines, and an increasing number of academic and research institutions now offer their own facility. A potential limitation of these is that a dataset is less likely to be found through serendipitous searches by individuals interested in a specific topic, in which case subject-specific repositories may be more appropriate. Repositories will have their own procedures and guidelines, and it is advisable to identify these in advance of posting. For example, there is often a process of user and project registration that needs to be navigated, which can take time.

Exactly what data are shared and how they are organized will vary across disciplines, but as a minimum one would expect to have access to a comprehensive, well-labeled data file that uses a non-proprietary software format and is accompanied by meta-data files for secondary users that describe complex data sheets or data extraction methods. This can take a substantial amount of time to collate, particularly if additional materials are required that would not otherwise have been produced. However, producing these while analysis is ongoing, rather than at the point of posting data, can significantly reduce the burden. We have found that these procedures have generated further benefits: data files are better organized and clearer, we have detailed records of analysis and extraction methods, and there is much better coherence across the group in the way we organize and archive our data. Focusing on the needs of secondary users has substantially enhanced the clarity, detail and effectiveness of research materials.

Open Access

The final aspect of the Open Science model, and one that is increasingly mandated by funding agencies, is making resulting publications Open Access. There are two routes to Open Access: green and gold. The gold route involves paying a fee to the publisher, so that the publication is immediately available (typically under a Creative Commons license, such as CC-BY). In their commitment to open access, funders such as Research Councils UK (RCUK) have allocated block grants to UK academic institutions to contribute to these fees, and most other funding mechanisms support Open Access charges. In contrast, the green route is free and involves publication in the normal way, the researcher posts a copy of the publication (typically a pre-print of the accepted version of the manuscript prior to copy-editing and proofreading) on an online platform. Journals differ in the embargo period stipulated by the publisher before this can be done; publisher copyright policies and

rules regarding self-archiving for individual journals can be determined via the SHERPA/RoMEO website (<http://www.sherpa.ac.uk/romeo/>).

We are fortunate enough to have RCUK funding at present, which enables us to make use of our institutional block grant to make our publications Open Access via the gold route. However, our institution is also developing a pre-print repository to enable all research to be made Open Access via the green route, and other similar repositories are available. Given the costs associated with the gold route, and limited research funding available, it is likely that these repositories will become an increasingly important platform for making publications Open Access.

Conclusions

There are a number of pragmatic reasons to adopt an Open Science model, as well as a motivation to do so because much research activity is ultimately publicly funded. In our experience, another reason to do so is that it serves to harmonize procedures at every step of the research pipeline, and improve quality control procedures. If researchers know from the outset that their materials, data and publications will be publicly available, this serves as a strong motivation to reduce errors at every step [9]. Human error is inevitable in any endeavor, including scientific research, and adopting processes that serve to minimize this should therefore be welcome. However, moving towards an Open Science model can be a substantial undertaking, and may require changes in procedures and the use of a number of platforms to make materials, data and publications publicly available. It may require discussions with institutional ethics committees and research governance teams, and in some cases may not be appropriate (e.g., where there is a risk of participant identification where sensitive information is involved). However, our experience of attempting to adopt an Open Science model has been extremely positive and we believe that it will ultimately improve the quality of our work.

Acknowledgements

The authors are members of the UK Centre for Tobacco and Alcohol Studies, a UKCRC Public Health Research: Centre of Excellence. Funding from British Heart Foundation, Cancer Research UK, Economic and Social Research Council, Medical Research Council, and the National Institute for Health Research, under the auspices of the UK Clinical Research Collaboration, is gratefully acknowledged.

References

1. West, R., *Promoting greater transparency and accountability in clinical and behavioural research by routinely disclosing data and statistical commands*. J Clin Epidemiol, in press.
2. Poschl, U., *Multi-stage open peer review: scientific evaluation integrating the strengths of traditional peer review with the virtues of transparency and self-regulation*. Front Comput Neurosci, 2012. **6**: p. 33.
3. Chambers, C.D., *Registered reports: a new publishing initiative at Cortex*. Cortex, 2013. **49**(3): p. 609-10.
4. Chambers, C.D., et al., *Instead of "playing the game" it is time to change the rules: Registered Reports on AIMS Neuroscience and beyond*. AIMS Neuroscience, 2014. **1**(1): p. 4-17.
5. Munafo, M.R. and E. Strain, *Registered reports: a new submission format at Drug and Alcohol Dependence*. Drug Alcohol Depend, 2014. **137**: p. 1-2.
6. Stahl, C., *Toward Reproducible Research*. Experimental Psychology, 2014. **61**(1): p. 1-2.
7. Nosek, B.A. and D. Lakens, *A method to increase the credibility of published results*. Social Psychology, 2014. **45**(3): p. 137-141.
8. Bloom, T. *Data access for the open access literature: PLOS's data policy*. 2013; Available from: <http://www.plos.org/data-access-for-the-open-access-literature-ploss-data-policy/>.
9. Munafo, M., et al., *Scientific rigor and the art of motorcycle maintenance*. Nat Biotechnol, 2014. **32**(9): p. 871-3.